Carbon-Credit Monitoring and Prediction in Smart Factory using Explainable AI and Data Analytics

Cosmas Ifeanyi Nwakanma *, Martin Uwakwe ‡, Ikenna Uzoma Ajere §,
Ebuka Chinaechetam Nkoro †, Love Allen Chijioke Ahakonye †, and Dong-Seong Kim †
* ICT Convergence Research Center, Kumoh National Institute of Technology, Korea.
‡Applied Statistics and Decision Analytics, Western Illinois University, USA.
§ Information Technology, Federal University of Technology, Owerri, Nigeria.
† IT Convergence Engineering, Kumoh National Institute of Technology, Korea.

***Abstract*—Monitoring and prediction are critical to meet the carbon reduction objective. However, relying solely on monitoring is not sufficient for local smart factories. Solutions must be built to help factories monitor, analyze, and predict their carbon footprint. In this work, a carbon smart proposal is presented incorporating analytics, and explainable AI for efficient carbon emission monitoring and prediction**

***Index Terms*—Carbon credit, Explainable AI, Data analytics, climate change.**

I. INTRODUCTION

Carbon credit refers to the security (in the form of a certificate or tradable license) that grants the holder the permission to emit one ton of carbon dioxide (CO_2) or its equivalent in any other greenhouse gases (GHG) to the environment [1], [2]. It is one of the mechanisms devised for factories to mitigate, remove, or avoid CO_2 and other GHG emissions. It is based on the two-legged *Cap-and-Trade* model whereby factories buy more credits allowing them to emit when their emissions have reached the allowed threshold or sell the surplus credits to other companies when their emissions are below the threshold and invest the money in projects that reduce pollution [1]. The model also incentivizes companies to save money through cost-effective emission reduction schemes and is the most preferred emission trading mechanism because it leads to measurable and verifiable carbon emission reduction. Artificial intelligence (AI) offers a set of multipurpose tools and techniques for predicting carbon footprints [3], monitoring carbon removal and CO_2 leaks [4] as well as reducing carbon emissions. Thus, researchers and companies have applied AI to reduce carbon emissions through improvements in product quality, optimization of supply chains, optimization of heating and cooling systems, and prioritization of the use of clean electricity over fossil fuels [5], [6].

Motivation - Previous and existing works are yet to explore the role of data analytics and explainable AI (XAI) methods which can provide insights and transparency for predicting CO_2 threats and mitigation [7]. Ignoring XAI and data analytics is not sustainable in this age of massive data generation and the growing need to meet the ZERO carbon goal of 2050.

Contribution - This paper's contributions are three-fold:

1) Proposed an ensemble Bagged Tree classification and prediction helping to monitor carbon emission which is an important variable to determine carbon creditworthiness.
2) Exploratory data analytics of Carbon emission from factory plants across continents.
3) An exploration of the SHapley Additive exPlanations (SHAP) explainability to gain insights into the CO_2 emission and prediction.

II. RELATED WORK

Carbon credits represent removing one ton of CO_2 from the environment and offer a way for businesses to achieve GHG reduction goals without disrupting operations [1]. Carbon credits are part of more effective climate change strategies but do not eliminate emissions. The field of carbon credits has gained significant attention in recent years due to the growing emphasis on sustainability and carbon neutrality [1], [2]. As a component of carbon trading, it promotes emission reduction for financial gain. Concerns about actual emission reduction are raised by varying project effectiveness and transparency. According to reports, the carbon commodities market surged to $851 billion in 2021, with a projected growth of $22 trillion by 2050 [1], [2].

AI tools are vital in efficiently managing and tracking carbon credits, enabling carbon footprint monitoring and progress toward carbon neutrality. By investing in carbon credits and leveraging AI, companies can contribute to a low-carbon future, earn profits, and promote emissions reduction efforts in line with the zero emissions target for 2050. Aamir *et al* utilized data from China and South Asian nations between 2001 and 2020 to predict carbon emissions, considering factors from industrial, energy, and transportation sectors [8]. This empirical study allows an understanding of the combined impact of various factors on emissions, particularly greenhouse gases like carbon dioxide, which are the primary drivers of climate change.

Additionally, other studies have employed AI and machine learning to forecast global carbon emissions, considering the impact of the COVID-19 pandemic on temporary emission reductions [9]. Furthermore, researchers have explored the use of blockchain technology to track and trade carbon emissions, demonstrating its potential for carbon emission trading and monitoring [10], [11]. In the context of air quality prediction, machine learning models were employed to forecast the air quality index in Delhi, India, using various evaluation metrics [12]. These diverse studies collectively highlight the significance of AI and ML in addressing carbon emissions and climate change challenges.

Researchers have investigated the application of AI techniques to monitor, analyze, and predict carbon emissions. Helo *et al* [13] developed a cloud-based distributed manufacturing execution system based on distributed AI which shortens the product development lifecycle thus reducing the carbon footprint of the production process. Mardani *et al* [3] utilized the economic growth and energy consumption dataset of Group of 20 (G20) nations to develop an efficient multi-stage methodology for predicting CO_2 emissions using a self-organizing map clustering algorithm to cluster the data, artificial neural network, and adaptive neuro-fuzzy inference system to construct the prediction models and the singular value decomposition for dimensionality reduction. They found that the multi-stage approach could predict CO_2 emissions with a high level of accuracy based on the two indicators. In [14], machine learning algorithms were used to minimize the carbon footprint of concrete in the construction industry.

Some researchers [15], [16] opined that the use of AI to reduce carbon emissions also comes with certain challenges such as ethical risks, model training contribution to carbon footprint, and the need for gainful insights into carbon footprint data. Consequently, Cowls *et al* [15] suggested that devising strategies to limit carbon emissions may require data that reveal patterns of human behavior in use cases where privacy concerns may be relevant. Thus, we proposed a combined usage of the XAI and data analytics to achieve carbon-smart monitoring and prediction using a smart factory public data - Carbon Monitoring for Action (CARMA) [17].

III. METHODOLOGY

This section captures the framework and overall idea of the projects as shown in Fig. 1.

A. Dataset Description and Pre-processing

The CARMAv3.0 dataset by the Center for Global Development was used in this project [17]. The dataset contains information about the CO_2 emissions, electricity production, corporate ownership, and location of more than 60,000 power plants in over 200 countries. The dataset provided an estimated rate at which a plant emits CO_2 (per unit of electricity generated) distributed across various continents as shown in Fig. 2.

B. Data Analytics Tool and Approach

The ***tidyverse*** collection of open-source packages was used for the exploratory data analysis. The classification and XAI framework were carried out using the sci-kit learn library,

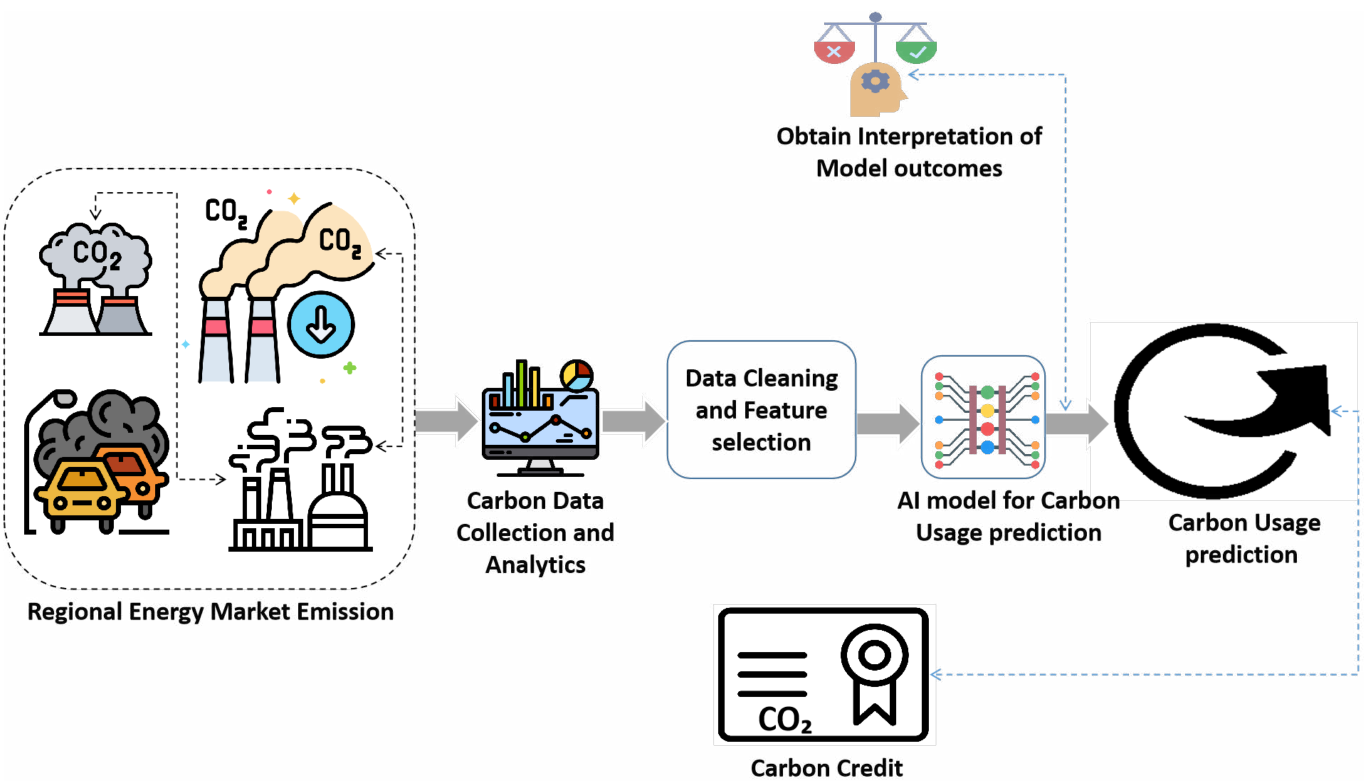

Fig. 1. Illustrating the stratified methodology for awarding carbon credits to holders: As depicted, the process entails the collection and analysis of regional carbon data, followed by the curation of pertinent features for model training. After training, post-hoc explainability techniques are harnessed to furnish lucid insights into model predictions, culminating in the issuance of a carbon credit certificate.

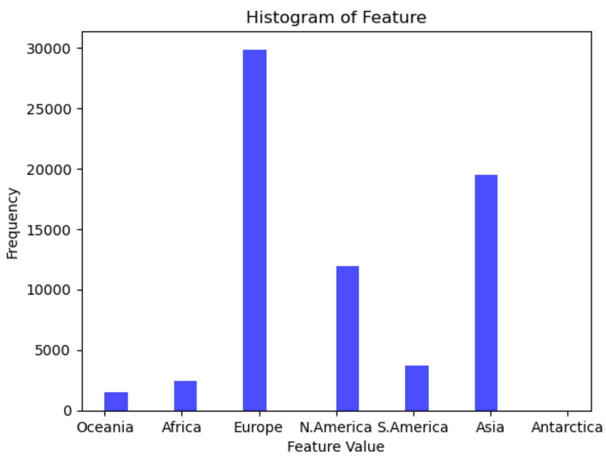

Fig. 2. The distribution of CARMA Dataset showing the Carbon Emissions across Continents. Europe has the highest emission recorded. Asia is next (It is argued that China is a major player). next are North America and South America.

which provides tree-based APIs. The experiment was run on Jupyter Notebook with an Intel(R) Core(TM) i3-7100 CPU @ 3.90GHz, and RAM 8.00 GB. After simulating various models, the best performance was achieved using the Ensemble Bagged Tree Classifier.

C. SHapley Additive exPlanations (SHAP) of Model

SHAP [18] allows for insights into the output or basis for the performance of a machine learning or deep learning model. It helps to understand the feature importance and how they contribute to the predictions' performance of machine learning or deep learning models.

IV. Performance EVALUATION

Prediction Model Diagnostics:

1) Residual standard error: This is the average amount that the response will deviate from the true regression line. It's 989500 in this model, which is quite large but needs to be interpreted in the context of the scale of carbon_present.
2) Multiple R^2: This value (0.4526) represents the proportion of variance in the dependent variable (carbon_present) that's explained by the independent variables. In this model, about 45.26% of the variability in carbon_present is explained by energy_present and intensity_present.
3) F-statistic and p-value: The very low p-value ($\leq$ 0.00000000000000022) suggests that the model is statistically significant.

Interpretation: Leveraging the CARMAv3.0 dataset, it was deduced that both energy_present and intensity_present are statistically significant predictors of carbon_present. About 45.26% of the variability in carbon emissions (in the present) is explained by the model. For each unit increase in energy generation in the present, the carbon emission increases by approximately 0.439850 units. Similarly, for each unit increase in intensity in the present, carbon emission increases by approximately 5.088198 units, assuming the other variable remains constant. The model is statistically significant based on the F-statistic and its associated p-value. Consequently, the Ensemble Bagged Tree outperformed at an accuracy of 90% higher than the linear regression prediction model.

A. Model Prediction Performance

The box plot presented in Fig. 3 shows that the prediction model reflected the true spread and outlier for the carbon future emissions.

Correlation Analysis The correlation analysis between carbon emission and other variables such as carbon intensity and energy production was investigated. Table I shows that, while there exists a strong correlation between energy production and carbon emissions **70%**, the relationship between carbon emission and intensity over time continues to be weak **2%**. Despite the weak relationship, the result shows it is a significant one. Table I summarizes these findings.

This indicates that plants that generate more energy also tend to emit more carbon. The strength of this relationship seems to be increasing slightly over time, suggesting that energy generation's impact on carbon emissions may be growing. As a caution, correlation does not imply causation. A causality test might be needed to investigate the role of other factors not included in the dataset.

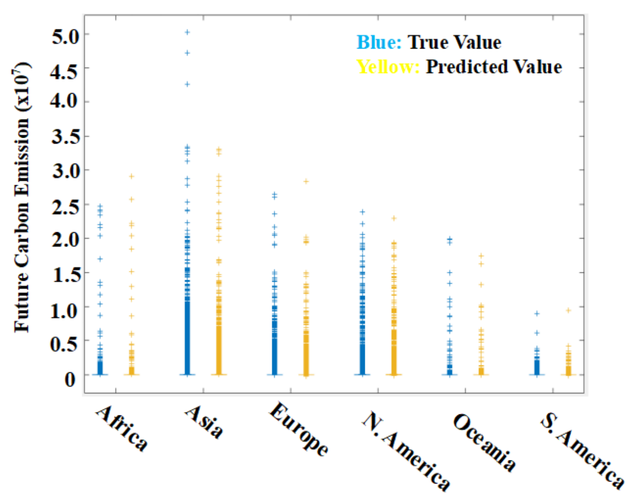

Fig. 3. Comparing the Model Prediction with the True value of the Future Carbon Emission according to Continents. This prediction is based on the former and present Carbon values in the CARMA Dataset. Our prediction when compared with the true value demonstrated above 90% accuracy.

TABLE I
CORRELATION ANALYSIS BETWEEN CARBON EMMISSION AND FACTORS SUCH AS PLANT ENERGY AND INTENSITY OF USAGE

Factors (variables)	Carbon_Past	Carbon_Present	Carbon_Future
Energy_Past	0.6848734		
Energy_Present		0.6725515	
Energy_Future			**0.7013609**
Intensity_Past	0.2552156		
Intensity_Present		0.0225209	
Intensity_Future			**0.02253723**

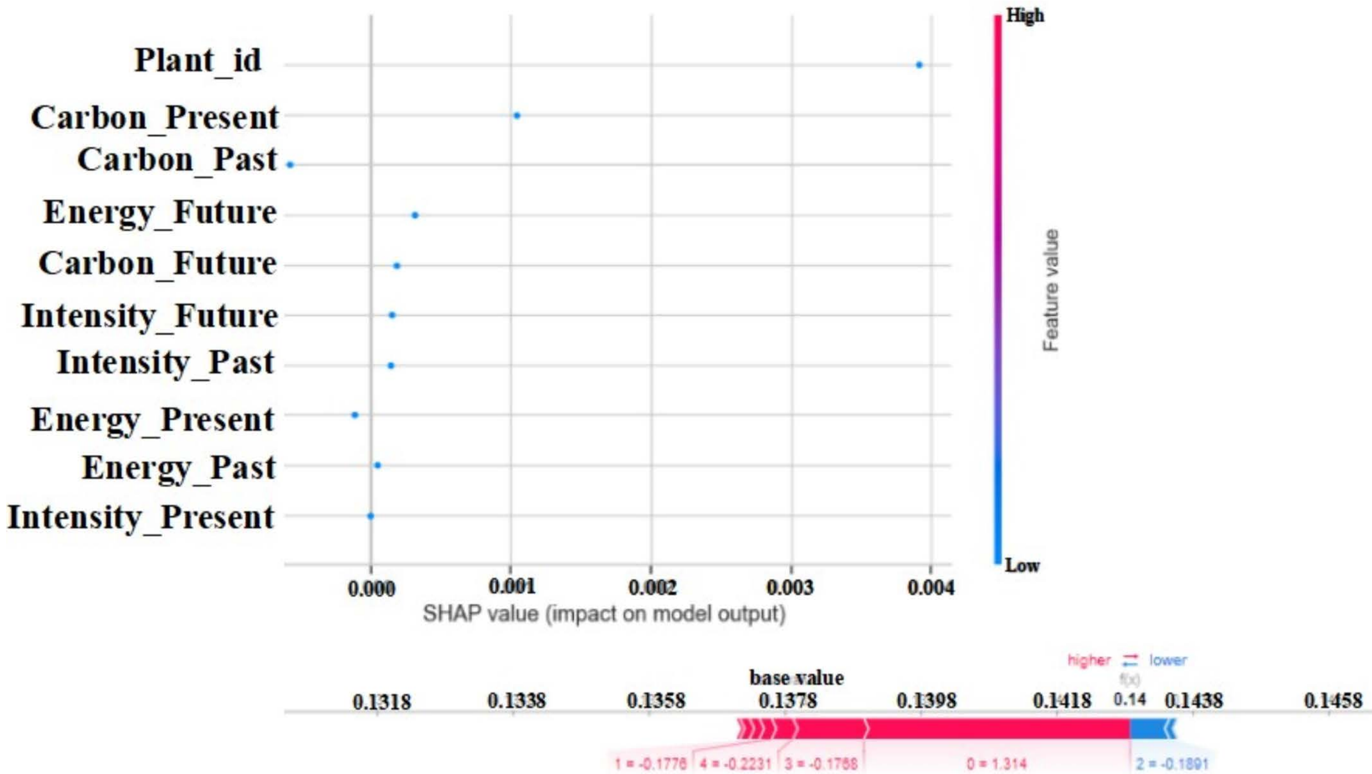

Fig. 4. SHAP ExplainabilityShowing the impact of the Energy Generation, the Intensity of plant usage on Carbon Emission in the Past, Present, and Future

B. Model Explainability with SHAP

The SHAP explainable result is presented in Fig. 4. The SHAP values show that there is a connection between the present and past carbon emissions by the plants (factories) and future emissions of carbon. Thus, leveraging the past and present carbon emission value of a smart factory can be used to determine the carbon credit to be awarded to such a plant. In addition, the SHAP values show that the energy usage is more reliable than the intensity of usage, thus forecasting the future energy a smart factory is likely to generate can give an insight into the extent of carbon to be emitted or conserved (in cases where the energy sources are alternative and clean sources that are carbon-smart).

V. CONCLUSION

In this work, we have leveraged data analytics for exploratory CO_2 future emission data and provided an explainable AI inquest into the relationship between factors impacting the emission of carbon. This insight has become critical if the Carbon credit objective goal for zero carbon in 2050 will be achieved. However, a major limitation of this research is the adoption of only one dataset due to the difficulty of accessing CO_2 data and the non-disclosure of carbon emissions. In the future, we hope to employ blockchain for an immutable carbon credit certificate.

ACKNOWLEDGMENT

This research was supported by the Priority Research Centers Program through the NRF funded by the MEST (2018R1A6A1A03024003), and by MSIT under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-2020-0-01612) supervised by the IITP.

REFERENCES

[1] J.-A. M. Lutfi, G. P. MacEwan, and J. D. Lauria-Banta, "AI and Carbon Credits: How the Emergence of AI Tools and Technologies Facilitates the Use of Carbon Credits," in *Energy Current Innovative Technology Insights*. FOLEY & LARDER LLP, February 10 2023.

[2] C. Blaufelder, C. Levy, P. Mannion, and D. Pinner, "A Blueprint for Scaling Voluntary Carbon Markets to Meet the Climate Challenge," in *McKinsey Sustainability*. McKinsey & Company, January 29 2021.

[3] A. Mardani, H. Liao, M. Nilashi, M. Alrasheedi, and F. Cavallaro, "A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques," *Journal of Cleaner Production*, vol. 275, p. 122942, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652620329875

[4] B. Chen, D. R. Harp, Y. Lin, E. H. Keating, and R. J. Pawar, "Geologic co2 sequestration monitoring design: A machine learning and uncertainty quantification based approach," *Applied Energy*, vol. 225, pp. 332–345, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918307372

[5] R. H. Kazi, T. Grossman, H. Cheong, A. Hashemi, and G. W. Fitzmaurice, "Dreamsketch: Early stage 3d design explorations with sketching and generative design," in *UIST '17: Proceedings of the 30th Annual ACM Symposium on User Interface Software and TechnologyOctober*, 2017, p. 401–414. [Online]. Available: https://doi.org/10.1145/3126594.3126662

[6] X. Zhang, G. Hug, J. Z. Kolter, and I. Harjunkoski, "Model predictive control of industrial loads and energy storage for demand response," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5. [Online]. Available: https://doi.org/10.1109/PESGM.2016.7741228

[7] L. Zhu, X. Zhou, W. Liu, and Z. Kong, "Total organic carbon content logging prediction based on machine learning: A brief review," *Energy Geoscience*, vol. 4, no. 2, p. 100098, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666759222000191

[8] M. Aamir, M. A. Bhatti, S. U. Bazai, S. Marjan, A. M. Mirza, A. Wahid, A. Hasnain, and U. A. Bhatti, "Predicting the Environmental Change of Carbon Emission Patterns in South Asia: A Deep Learning Approach Using BiLSTM," *Atmosphere*, vol. 13, no. 12, p. 2011, 2022.

[9] Y. Meng and H. Noman, "Predicting CO2 Emission Footprint Using AI through Machine Learning," *Atmosphere*, vol. 13, no. 11, p. 1871, 2022.

[10] D. Effah, B. Chunguang, F. Appiah, B. L. Y. Agbley, and M. Quayson, "Carbon Emission Monitoring and Credit Trading: The Blockchain and IOT Approach," in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2021, pp. 106–109.

[11] P. Yuan, X. Xiong, L. Lei, and K. Zheng, "Design and Implementation on Hyperledger-Based Emission Trading System," *IEEE Access*, vol. 7, pp. 6109–6116, 2018.

[12] R. Kumar, P. Kumar, and Y. Kumar, "Time Series Data Prediction using IoT and Machine Learning Technique," *Procedia computer science*, vol. 167, pp. 373–381, 2020.

[13] P. Helo, M. Suorsa, Y. Hao, and P. Anussornnitisarn, "Toward a cloud-based manufacturing execution system for distributed manufacturing," *Computers in Industry*, vol. 65, no. 4, pp. 646–656, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361514000311

[14] P. Thilakarathna, S. Seo, K. K. Baduge, H. Lee, P. Mendis, and G. Foliente, "Embodied carbon analysis and benchmarking emissions of high and ultra-high strength concrete using machine learning algorithms," *Journal of Cleaner Production*, vol. 262, p. 121281, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652620313287

[15] J. Cowls, A. Tsamados, M. Taddeo *et al.*, "The ai gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations." *AI & Soc*, vol. 38, p. 283–307, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00146-021-01294-x

[16] R. Nishant, M. Kennedy, and J. Corbett, "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda," *International Journal of Information Management*, vol. 53, p. 102104, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401220300967

[17] K. Ummel, "Carma revisited: An updated database of carbon dioxide emissions from power plants worldwide," in *Carbon Monitoring for Action (CARMA) Database*, 2012. [Online]. Available: https://opennetzero.org/dataset/carbon-monitoring-for-action-carma-database

[18] C. I. Nwakanma, L. A. C. Ahakonye, J. N. Njoku, J. C. Odirichukwu, S. A. Okolie, C. Uzondu, C. C. Ndubuisi Nweke, and D.-S. Kim, "Explainable artificial intelligence (xai) for intrusion detection and mitigation in intelligent connected vehicles: A review," *Applied Sciences*, vol. 13, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/3/1252